Data Mining of Building Electrical Information Based on Radial Basis Function Neural Network

Norman C. F. Tse[1], Wing W. Y. Ng[2], T. T. Chow[1], John Chan[3], L L Lai[3], Daniel S. Yeung[2], Jincheng Li[2]
[1]City University of Hong Kong, HKSAR, China
[2]School of Computer Science and Engineering, South China University of Technology, China
[3]City University London, UK
bsnorman@cityu.edu.hk

***Abstract*--This paper presents a neural network algorithm for data mining in building LV electrical power information. The power information is recorded by web-based power quality monitoring system. Power information is recorded continuously and stored in a central server system. Presently events were identified by power engineers but in the prototype, an expert system will be used to identify events instead. Neural network approach based on the Radial Basis Function Neural Network (RBFNN) was developed to predict power events in the building LV electrical network. The approach provides useful information for facility managers to conduct planning and operation. The proposed algorithm was tested with power data of a commercial building in Hong Kong. The prediction result by using one week of data achieved 75% accuracy. Further works would be conducted to test the algorithm with more data.**

Index Terms*—*PQ monitoring, building LV electrical network, micro-grid, data mining, neural network.

I. INTRODUCTION

Over 60% of electricity in Hong Kong is consumed in buildings. Electricity generation in Hong Kong is mainly by burning coals, oil and natural gas. They are the main cause of air pollution and green house effect. Researchers and engineers on one hand explore the potentials of alternative renewable energy resources for electricity generation, on the other hand it is believed that energy saving would be the most effective measure in reducing harm to the environment in short run. Energy efficiency and conservation thus attracts more and more attention from government officials, facility managers and the general public. Measures and guidelines have been proposed in many countries to reduce energy consumption and wastage. As the major energy usage in buildings takes the form of electricity, the need to monitor the electricity usage is evident and this constitutes an important component in the whole energy conservation and efficiency campaign. The intelligent power monitoring and diagnosis system (IPMDS) [1, 2] has been employed in electrical power system transmission and distribution for many years. They are dedicated system employing accurate and expensive measuring devices, networked together to provide continuous surveillance of the power system. However more and more buildings in Hong Kong employ the intelligent power monitoring and diagnosis system (IPMDS) for power quality and energy efficiency monitoring purposes. They are not only for standard alone applications, but also being integrated with building automation system for energy audit purpose and fault reporting. The rapid changes that have taken place in the electricity industry make it of vital importance that building operators have easy access to process data and to obtain information on electricity usage. Power information is an important quantity to manage in legal terms as well as in money terms.

The development of mirco-grid systems [3] in recently years has also called for an advancement in intelligent network power monitoring system design.

A powerful data mining algorithm is necessary to extract useful information from the power data recorded and stored by IPDMS. The algorithm should perform many functions: classification of loads, characterization of events, faults detections, trends deduction and prediction. This paper introduces a data mining algorithm based on Radial Basis Function Neural Network (RBFNN) to extract and consolidate power information obtained from IPDMS.

The paper is arranged this way: Section II describes the basic construction and functions of a typical web-based building intelligent power monitoring and diagnosis system; Section III discusses the common trending and event reporting functions performed by the IPMDS; Section IV describes a localized generalization error model for the RBFNN for power information extraction and prediction; Section V presents a preliminary study on applying the RBFNN in predicting power events; and Section VI presents the conclusions and further works.

II. INTELLIGENT POWER MONITORING AND DIAGNOSIS SYSTEM (IPMDS)

A typical LVMDM system commonly installed in commercial building in Hong Kong is shown in Fig.1 [4]. The system consists of the following basic hardware components:

- power quality monitoring instrument;
- communication network; and
- central server and workstations.

Normally the IPMDS is web-based; users are able to view real-time information over the webpage. Other communication technologies are being investigated too. Some of the popular

approaches are power line communication, Zigbee wireless protocol, GPRS, etc.
The monitoring functions of the IPMDS can be grouped into four major categories [5]:

- real time power information (rms values of phase & line voltages, rms values of phase currents and the neutral current, and total power factors);
- energy data (kWh, KVA, KW and kVAr);
- event data (minimum /maximum values with time stamp for the (1) and (2) above, sag/swell voltage or set-point current trigger);
- power harmonics (THD of voltage and current harmonics; and individual odd current harmonic measurement).

The recording interval of the system is adjustable. A 15-min reporting interval is usually adopted to reduce memory size and to ease useful information retrieval.

All the information recorded by the power quality monitoring instrument will be stored in the central server and workstation as a power database. The power database is subdivided into several database sections to perform the following functions:

- trending function;
- event reporting function; and
- compliance reporting.

Trending function is used to store data on voltage, current, frequency and flickering, power factor, harmonics and power usage. Events reporting can be generated on regular basis. In addition to the useful functions of providing trending and event reports, the system would provide automatic real-time fault reporting functions. Depending on the type of alarms required, the system would generate alarm messages such as sag/swell duration, time of the event happening, site location, to the clients/users through email and pager. Other trending and reporting functions can be derived to meet operating needs of buildings.

III. TRENDING AND EVENTS REPORTING FUNCTIONS

The functions are derived to assure compliance to statutory requirements, to conduct long term power quality monitoring for determining preventive practices, and to establish a power quality and energy efficiency profile. Continuous tracking of power consumption and electrical quantities provide useful data for planning future plant expansion and for ensuring that the existing distribution system is adequate. It is also useful for planning new alternative power supply system, such as wind energy generation systems and photovoltaic cells. With mandatory electrical energy efficiency requirements implemented in many countries, the trending and events report functions would provide useful information for energy efficiency audit and compliance checking.

Event reporting and compliance reporting are normally performed by an expert system. Rules used are derived from statutory requirements related to power quality and the operational needs of the facility management.

IV. LOCALIZED GENERALIZATION ERROR MODEL FOR RBFNN

The Radial Basis Function Neural Network (RBFNN) is adopted to model the input-output mapping of the given power event prediction problem. A RBFNN is defined as:

$$f(\mathbf{x}) = \sum_{j=1}^{M} w_j \exp\left(\left\|\mathbf{x} - \mathbf{u}_j\right\|^2 / \left(-2v_j^2\right)\right) \qquad (1)$$

where x, u_j, w_j, v_j and M denote the input sample, the vector of the center of the j^{th} hidden neuron in the RBFNN, the connection weight between the output layer and the j^{th} hidden neuron, the cluster width of the j^{th} hidden neuron and the number of hidden neurons, respectively. However, solving a power event prediction problem with neural network requires human expert to select the network architecture by experience. Neural networks built in this way may not be optimal. The ultimate goal of solving machine learning problems is to minimize the generalization error of a classifier for future unseen samples. Therefore, the Localized Generalization Error Model (L-GEM) for RBFNN was proposed in [6, 7] to provide an analytical upper bound of the generalization error of a trained RBFNN.

For a given power event prediction problem, one wants to build an RBFNN with minimal generalization error for future data. Therefore, the network architecture yielding the smallest L-GEM value will be selected. The L-GEM provides the upper bound of Mean-Square-Error (MSE) of unseen samples similar to the training samples. This is counterproductive to model the error of very dissimilar unseen samples because one may not expect any neural network to recognize them correctly and this large error may not be helpful in architecture selection. With probability $(1-\eta)$, the L-GEM value of an RBFNN (R_{SM}^*) is given by:

$$R_{SM}^* = \left(\sqrt{\frac{1}{3}Q^2\sum_{j=1}^{M}\upsilon_j + \frac{0.2}{9}Q^4 n\sum_{j=1}^{M}\zeta_j} + \sqrt{R_{emp}} + 1\right)^2 + \varepsilon \qquad (2)$$

where,

$\varepsilon = B\sqrt{\ln\eta / (-2N)}$,

$E(s_j) = \sum_{i=1}^{n}\left(\sigma_{x_i}^2 + \left(\mu_{x_i} - u_{ji}\right)^2\right)$,

$$\upsilon_j = \varphi_j \left(\sum_{i=1}^{n} \left(\sigma_{x_i}^2 + \left(\mu_{x_i} - u_{ji} \right)^2 \right) \Big/ v_j^4 \right),$$

$$s_j = \left\| \mathbf{x} - \mathbf{u}_j \right\|^2, \ \zeta_j = \varphi_j / v_j^4,$$

$$\varphi_j = \left(w_j \right)^2 \exp\left(\left(Var\left(s_j \right) / 2v_j^4 \right) - \left(E\left(s_j \right) / v_j^2 \right) \right),$$

$$Var\left(s_j \right) = \sum_{i=1}^{n} \left(\begin{array}{l} E_D\left[\left(x_i - \mu_{x_i} \right)^4 \right] - \left(\sigma_{x_i}^2 \right)^2 \\ + 4E_D\left[\left(x_i - \mu_{x_i} \right)^3 \right] \left(\mu_{x_i} - u_{ji} \right) + 4\sigma_{x_i}^2 \left(\mu_{x_i} - u_{ji} \right)^2 \end{array} \right),$$

and Q, n, N, B and R_{emp} stand for the maximum difference between the training samples and unseen samples in each input feature, the number of input features, the number of training samples, the maximum value of the Mean-Square-Error which could be pre-selected and the training error of the RBFNN.

The algorithm of selecting optimal RBFNN architecture is as follows:

- start with M = 1
- find the centers of RBFNN using K-Means with M clusters
- find the widths of RBFNN by computing the minimum distance between centers
- compute the connection weights in RBFNN by a least-square method with a least square method
- compute the L-GEM value R^*_{SM}
- if $M \leq N$, go to Step 2 with M = M + 1
- output the RBFNN with smallest R^*_{SM} value

The dataset for a given power event prediction problem is divided into training and testing datasets. The training dataset will be used in the training of RBFNN while the testing will not be involved in the training process. The Testing dataset serves as unseen future data of the problem and is used to examine the generalization capability of the trained RBFNN.

V. PRELIMINARY STUDY RESULTS

The power information dataset used for the study was recorded by an industrial grade power quality meter. One week of electrical data is collected at 15-min interval of a power distribution circuit in a commercial building in Hong Kong. The circuit monitored is rated as a 3-phase, 3-wire system, with a nominal line voltage of 380 volt, 50 Hz. The following electrical data is captured:

- phase voltages;
- phase currents;
- voltage harmonics;
- current harmonics;
- supply voltage frequency;
- total and per phase power (kWh, KW, KVA, KVAr, TPF).

An expert system will be implemented in future to conduct events reporting according to the following parameters:

- frequency variation should not be more than ±1%
- voltage imbalance among the phase voltages should not be more than ±3%
- voltage variations should not be more than ±10%
- voltage change between any two consecutive data should not be more than 10%
- voltage value smaller than or equal to 10% of the nominal voltage should be classified as a voltage interruption
- voltage harmonics should not be more than 5%
- total power factor should not be lower than 0.85
- current magnitude change of more than 10% of the average current in the previous hour
- current imbalance among the three phases should not be more than 10% of the average phase currents at any instant;
- current harmonics at any instant should not be more than the following requirements:
- phase current < 40A, current THD ≤ 20%
- 40A ≤ phase current < 400A, current THD ≤ 15%
- 400A ≤ phase current < 800A, current THD ≤ 12%
- 800A ≤ phase current < 2000A, current THD ≤ 8%
- phase current ≥ 2000A, current THD ≤ 5%
- total real power change of more than 10% of the average total real power in the previous hour.

Presently, a power engineer will identify all the events. The start and end time of events were reported. In this study, all events identified are related to current imbalance and current changes.

The RBFNN is then applied to the dataset for predicting future events. Half of the dataset was used for training and the remaining half was used for testing. The training dataset is divided into two sets: dataset without any events and dataset with events; they are clustered correspondingly for feature representation and then RBFNN was applied. The results show that a prediction accuracy of 75% was achieved.

VI. CONCLUSIONS

The paper presents a preliminary study on applying the RBFNN to power events predictions. This is the first time that RBFNN applied to this problem with such learning algorithm. The results show that the RBFNN is able to predict events at an acceptable accuracy. This would be a very useful tool for facility managers in planning and operation. It is expected that the prediction accuracy would be greatly improved when more data is used in the training process.

ACKNOWLEDGEMENT

The work described in this paper was fully supported by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China [Project No. RGC Ref. No. CityU 121008].

REFERENCES

[1] Mark McGranaghan, "Trends in Power Quality Monitoring", IEEE Power Engineering Review, Oct. 2001, pp. 3-9, 21.
[2] Afroz K. Khan, "Monitoring Power for the Future", Power Engineering Journal, April 2001, IEE, pp. 81-85.
[3] MICROGRIDS – Large Scale Integration of Micro-Generation to Low Voltage Grids", EU Contract ENK5-CT-2002-00610, http://microgrids.power.ece.ntua.gr.
[4] Lee R. P. K., Lai L. L., Tse Norman. "A web-based multi-channel power quality monitoring system for a large network". Proceedings of 5th International Conference on Power System Management and Control, IEE UK, London, 17-19 April 2002. pp. 112-117.
[5] Lee Ringo P. K., Tse Norman, Ki Simon, "Application of digitized monitoring and diagnosis system for building low-voltage power distribution system," Proceedings of the 3rd Power Symposium 2004, IEE Hong Kong, China, June 2004, pp. 2/1 to 2/5.
[6] Daniel S. Yeung, Wing W. Y. Ng, Defeng Wang, Eric C. C. Tsang, Xi-Zhao Wang, "Localized Generalization Error and Its Application to Architecture Selection for Radial Basis Function Neural Network", IEEE Transaction on Neural Network, 2007, pp. 1294 – 1305.
[7] Wing W. Y. Ng, Andres Dorado, Daniel S. Yeung, Witold Pedryca. Ebroul Izquierdo, "Image Classification with the Use of Radial Basis Function Neural Networks and the Minimization of the Localized Generalization Error", Pattern Recognition, 2007, pp. 19 – 32.

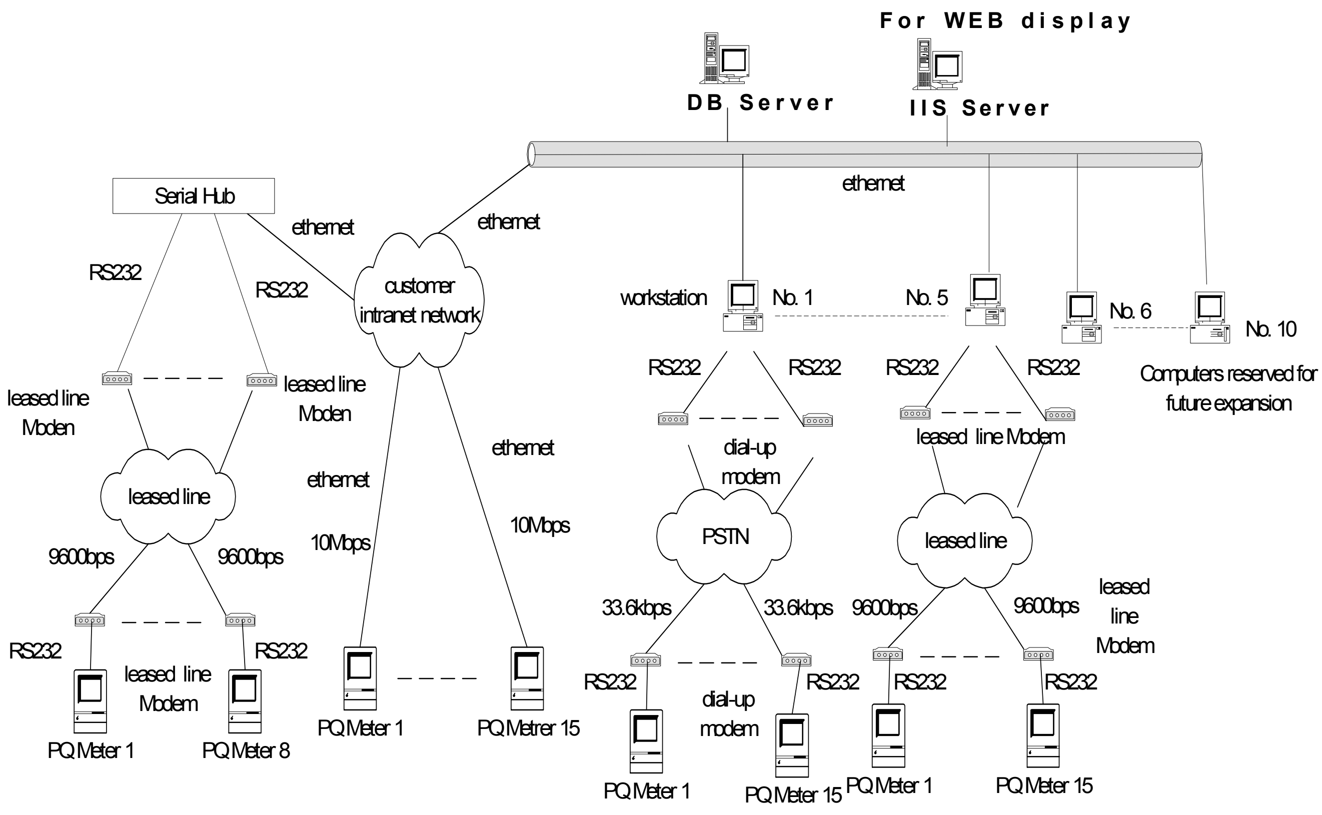

Figure 1. Power Quality System Network Layout.(extracted from [4])